

Article

Privacy-Preserving Electrocardiogram Monitoring for Intelligent Arrhythmia Detection [†]

Junggab Son ¹, Juyoung Park ², Heekuck Oh ³, Md Zakirul Alam Bhuiyan ⁴, Junbeom Hur ^{5,*} and Kyungtae Kang ^{3,*}

¹ Department of Computer Science, Kennesaw State University, Marietta, GA 30060, USA; json@kennesaw.edu

² Sustainable Management Strategy, Korea Expressway Corporation, Gimcheon 39660, Korea; juyoung@ex.co.kr

³ Department of Computer Science and Engineering, Hanyang University, Ansan 15588, Korea; hkoh@hanyang.ac.kr

⁴ Department of Computer and Information Sciences, Fordham University, Bronx, NY 10458, USA; mbhuiyan3@fordham.edu

⁵ Department of Computer Science and Engineering, Korea University, Seoul 02841, Korea

* Correspondence: jbhur@korea.ac.kr (J.H.); ktkang@hanyang.ac.kr (K.K.); Tel.: +82-31-400-5235 (K.K.)

† This paper is an extended version of our paper published in Park, J.; Lee, K.; Kang, K. Intelligent Electrocardiogram Monitoring System for Early Arrhythmia Detection. In Proceedings of the IEEE International Conference on Advanced Information Networking and Applications (AINA 2014), Victoria, BC, Canada, 13–16 June 2016.

Received: 11 April 2017; Accepted: 7 June 2017; Published: 12 June 2017

Abstract: Long-term electrocardiogram (ECG) monitoring, as a representative application of cyber-physical systems, facilitates the early detection of arrhythmia. A considerable number of previous studies has explored monitoring techniques and the automated analysis of sensing data. However, ensuring patient privacy or confidentiality has not been a primary concern in ECG monitoring. First, we propose an intelligent heart monitoring system, which involves a patient-worn ECG sensor (e.g., a smartphone) and a remote monitoring station, as well as a decision support server that interconnects these components. The decision support server analyzes the heart activity, using the Pan-Tompkins algorithm to detect heartbeats and a decision tree to classify them. Our system protects sensing data and user privacy, which is an essential attribute of dependability, by adopting signal scrambling and anonymous identity schemes. We also employ a public key cryptosystem to enable secure communication between the entities. Simulations using data from the MIT-BIH arrhythmia database demonstrate that our system achieves a 95.74% success rate in heartbeat detection and almost a 96.63% accuracy in heartbeat classification, while successfully preserving privacy and securing communications among the involved entities.

Keywords: body sensor networks; biomedical computing; electrocardiography; arrhythmia detection; communication system security; privacy of patients

1. Introduction

Cyber physical systems have emerged as a promising paradigm for enriching the interactions between physical and cybernetic components. Recent advances in sensing technology and smart devices, which are the most important devices facilitating cyber-physical systems, have drastically altered the shape of the current healthcare environment, while presenting numerous opportunities and challenges in patient monitoring and assistance. This novel paradigm enables patients to monitor their physical conditions using smart devices [1], which has been particularly useful for chronic diseases that can become life threatening, such as high blood pressure, hypernatremia and various

heart diseases that are common globally [2]. By monitoring a chronic disease, a patient can deal with this transformation at an early stage. In particular, monitoring arrhythmias is of high importance, because they are an extremely common initial symptom of cardiac arrest or myocardial infarction.

Arrhythmia can be detected by analyzing an electrocardiogram (ECG), which measures the rate and regularity of heartbeats and is characterized by five peaks and valleys labeled P, Q, R, S and T (see Figure 1). The amplitude and duration of the P-Q-R-S-T wave provides information regarding the heart disease. The position and distance of the PR interval and segment, the ST interval and segment and the QT interval and QRS complex can be used in a diagnosis [3].

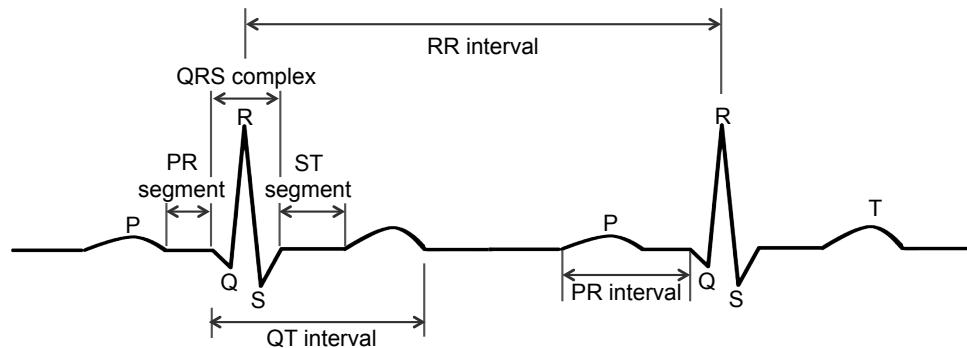


Figure 1. Structure of ECG.

Several successful ECG monitoring systems have been developed [2–6]. These systems possess many advantages, such as allowing for ambulatory patient care, resulting in safer and more affordable healthcare. However, ECGs are primarily interpreted by medical experts, and most patients themselves cannot obtain the information firsthand. Thus, considerable attention has recently been devoted to the computerized automatic analysis of heart activity, which includes both the detection of the heartbeat in an electrocardiogram and the classification of its type. Recent advances in portable devices, such as smartphones and tablet PCs, also make it possible for users to more easily self-monitor their ECG status and classification results [4].

In order to exploit this potential, we propose a novel and intelligent ECG monitoring system (I-ECG) that analyzes and interprets heart activity by automatically detecting and classifying heartbeats based on the Pan–Tompkins [7] and C4.5 algorithms [8–10]. Although this was developed to be effective, simple to use, sustainable and reliable, it cannot be directly applied in real-world applications, owing to the following security and privacy problems. First, individuals close to the patients can illegally obtain the ECG signal during the communication between the sensors and the smartphone. Second, a decision support server (DSS) can manipulate collected ECG signals, as well as the status of a particular user. The health information that can be obtained through these security flaws is sensitive personal data, and thus, it could be abused for monetary purposes. In addition, the privacy of patients could be invaded by the abuse of this information.

It should be noted that many articles treat the privacy of patients as a primary concern in the usage of personal healthcare systems [11–13], and some of these assert that patient privacy should be preserved at all times by law [12]. Therefore, healthcare systems must be designed with consideration given to the security, in order to provide justifiably trusted services. Simultaneously, the system should be designed to safeguard the privacy of patients.

Contributions: The main objective of this study is to develop a privacy-preserving I-ECG that detects arrhythmia in the early stage. The contributions can be summarized as follows:

- We outline the guidelines for an intelligent ECG monitoring system and service model, and propose an algorithm for automatic detection and classification of arrhythmia, which realize intelligent monitoring as a decision support system.

- Additional computations performed by the body sensors to enable encryption of ECG data to protect it from attackers generate heat on the sensors, which can result in discomfort for the patient. To address this problem, we have developed an efficient and effective solution for encryption of ECG data in the body sensors. This solution selects the cipher with minimal overhead among the ciphers considered as candidates for our environment. In addition, our solution protects the patients' privacy against attackers, including protection of data at the DSS, as well as monitoring stations by adopting changeable pseudonyms.
- We simulated our detection and classification algorithm for decision support using the MIT-BIH arrhythmia database [14], and the simulation results demonstrate that the DSS achieved an overall accuracy of 95.74%, with a sensitivity of 97.21% and a specificity of 94.26% for heartbeat detection. The DSS also achieved an overall accuracy of 96.63%, with a sensitivity of 95.44% and a specificity of 97.82% for heartbeat classification including privacy preservation measures.
- We demonstrate that our scheme exhibits a high efficiency compared with conventional cryptographic algorithms and an enhanced robustness against inside attackers who have access to the DSS, as well as against outside attackers who have the ability to eavesdrop on data from wireless communications.

The remainder of this paper is organized as follows. In Section 2, we briefly review related work regarding ECG monitoring systems and applications, with the focus on security and privacy. In Section 3, we describe the system models, threat models and security objectives. We describe our primary contribution, which is the privacy-preserving ECG monitoring system, in Section 4. In Section 5, we outline our system implementation. In Section 6, we evaluate and simulate our proposed system. Finally, we provide our conclusions in Section 7.

2. Related Work

Since the 1970s, researchers worldwide have been developing diagnostic systems that enable patients to make ECG recordings at home and transfer data to a cardiologist [2–6]. Research regarding remote ECG monitoring systems continues today, with smartphones being employed. Essentially, remote ECG monitoring systems [5,15,16] require only a simple architecture between the sensor device and a smartphone. The system extracts ECG signals through a lead cable, transmits the signal data via Bluetooth and then processes and displays the ECG waveform on a personal computer [15] or smartphone [5,16]. The system displays the ECG waveform on the device for remote monitoring. However, it does not transmit the ECG waveform to a medical expert. Therefore, the patient effectively receives no real-time medical service.

Systems on a healthcare server expand the fundamental remote monitoring system by offering an option to store and access ECG data. With the increasing popularity of Internet access through mobile phones, these systems provide an ideal platform between remote monitoring systems and patients. To display the ECG, many studies have employed a web component through a PC [17,18], single-chip microcomputer [19] or smartphone [2,18,20].

One proposed design for a remote monitoring “tele-medicine” system and web server consists of the client side, a general packet radio service (GPRS) modem and the server side [18]. The client side could be interpreted as the combination of the ECG, collection equipment, user interface and microprocessor. The GPRS modem, which is used to transmit the ECG signals, provides a large geographical mobility coverage range. The server side is divided into the back and front ends. The server front end is used to display the ECG signal on the web for patients and doctors, and the back end is designed to receive data from the GPRS modem and store it in a database. However, such designs are inadequate for medical experts, because they do not provide the information required to analyze the ECG signals.

In the recent past, many studies have applied QRS detection algorithms to healthcare servers. Such systems can record ECG signals on a web server and facilitate their analysis. One prototype uses a mobile phone as a gateway for transmitting measured ECG data back to the medical cloud using

3G mobile telecommunications or WiFi [6]. The system can also calculate the beats of the heart as RR intervals, which is the time between consecutive R-waves, on the mobile phone. Another system offers not only QRS detection, but also a priority-based alarm messaging service [4].

Although these approaches facilitate the convenient measurement and analysis of body status, they still encounter problems with security and privacy. As a number of healthcare applications deal with physical information, which constitutes sensitive personal information, data leakage and invasions of privacy in health monitoring systems are significant issues [11–13]. In 2012, Ma et al. proposed a simple, but effective security solution for ECG signals based on an ECG compression algorithm [21]. Regarding security and privacy challenges in mobile healthcare, Lu proposed an efficient user-centric privacy access scheme based on attribute-based access control and a novel privacy-preserving scalar product computation [22]. This scheme employs a body sensor node that can monitor various types of health information, and it has sufficient resources to apply a widely-employed symmetric encryption scheme such as AES. On the other hand, in this work, we focus on a body sensing device that only monitors an ECG, and thus, a smaller sensing device can be employed, which is more comfortable for patients. Because patients wear the sensing device for the entire day, it is an advantage to use a smaller device. Regarding this aspect, the scheme described above based on a symmetric key and hash functions requires more circuits and larger devices to handle the required cryptographic functions. Therefore, it remains a challenge to design a secure and privacy-preserving scheme for resource-constrained remote healthcare monitoring systems that utilize body sensors of a reduced size.

More recently, there have been attempts to deal with the use of ECG for identity recognition and biometric authentication [23,24]. For instance, Peter et al. [23] described the design and implementation steps required to realize an ECG-based authentication system in body area sensor networks and utilized ECG features for this purpose. Tan et al. [24] focused rather on enhancing the effectiveness and robustness of a biometric recognition system using a combination of random forest and wavelet distance measure classifiers. The ECG application investigated in this study is slightly different and has different design objectives. It is more oriented towards the secure transmission of ECG data to remote servers for analysis and to the application of machine learning to arrhythmia recognition. Enhancing the extent to which arrhythmia can be accurately detected is one of the eventual objectives of our study, and the energy is not our primary concern, because decisions are made on a remote server. Regarding networks, we are considerably interested in secure delivery over wide-area networks, rather than data transmission over body-area networks.

3. System Models, Threat Models and Security Objectives

This section describes the system models, threat models and security objectives of our system.

3.1. System Model

We are particularly interested in recording and accumulating ECG data for each patient over a long period of time, performing supervised learning based on the key ECG features and intelligent heartbeat classification for early and automatic detection of arrhythmia (we used a classifier based on a decision tree, which is constructed using the C4.5 learning algorithm). The crucial design consideration is the level of accuracy with which arrhythmia can be recognized. To achieve the desired accuracy, it is imperative that all of the acquired ECG data should be stored in the permanent storage of the DSS (i.e., database) and not on mobile devices with limited storage space.

Our system comprises four components: a revocation authority, an ECG sensing entity, a decision support server (DSS) and a monitoring station, as depicted in Figure 2. The patient is able to lead a normal life while the sensor continuously acquires ECG data and sends them to a Bluetooth-enabled smartphone. The smartphone relays the data to the DSS, which associates incoming data with the records for the patient in a database and analyzes them. The heartbeat is detected using the Pan-Tompkins algorithm, which has been shown to be an effective QRS detection scheme [25], and then classified using a decision tree. If the system identifies congestive heart failure or an irregular

heartbeat, then an alarm is sent over the Internet connection to a monitoring station, where medical personnel can carry out appropriate actions.

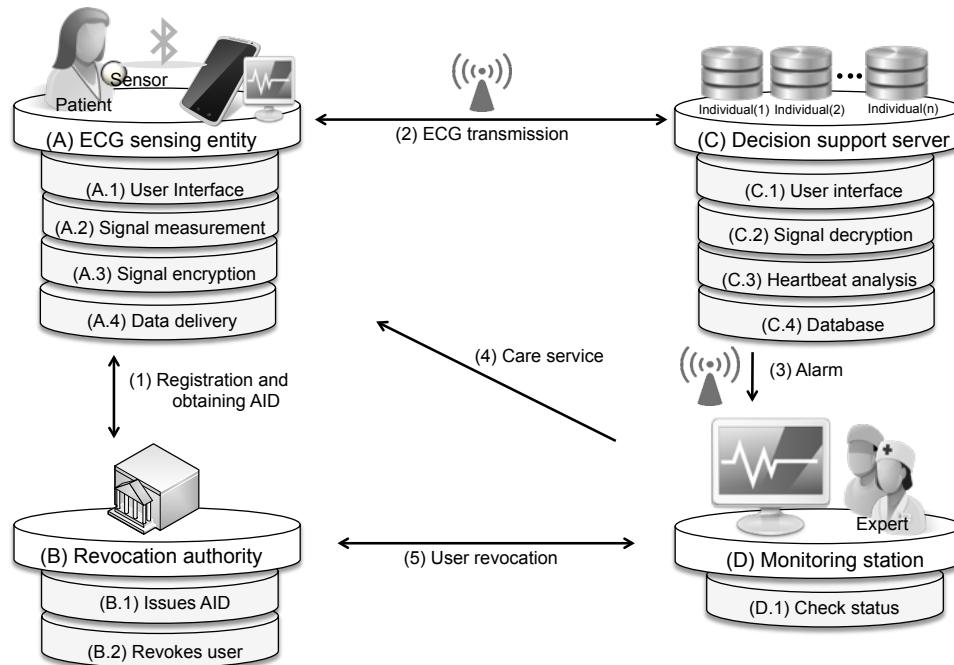


Figure 2. Overview of the proposed system architecture. AID, anonymous ID.

3.1.1. Revocation Authority

We assume the existence of a revocation authority (RA) to hide a user's identity, while being able to make it available during an emergency. The RA has three roles in the proposed scheme. First, it manages each user's identity using a related anonymous ID (AID). Second, it reveals a user's identity and contact information in an emergency, following a request by the monitoring station. Third, it issues certification for each entity's public key. As the RA plays an important role in user identity management and public key confirmation, we treat the RA as a trusted third party. A government agency could set up and manage the RA for public welfare purposes.

3.1.2. ECG Sensing Entity

The smartphone of a patient collects ECG signals every day. Figure 3 illustrates the process of signal measurement and delivery. A sensor node is attached to the patient's body. The signal from the sensor node is converted to a digital value using an analog-to-digital (ADC) converter and then sent to a microcontroller unit (MCU) through a serial peripheral interface, which is used primarily to communicate between chips. The signal at the MCU is sent to a Bluetooth module using a universal asynchronous receiver and transmitter. Subjects can view their ECG through a graphical interface designed for their mobile device. Although it is possible to perform self-monitoring, the ECG signal should also be transmitted to a DSS, because a typical layman cannot independently interpret the signal. Accordingly, the signal on the smart device is transmitted over a wireless network.

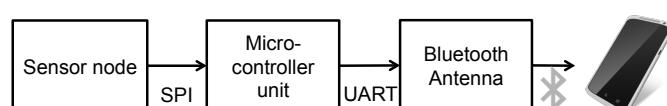


Figure 3. ECG signal measurement and delivery.

3.1.3. Decision Support Server

The DSS analyzes multiple aspects of the ECG. As a convenience for medical experts, it acts as a monitoring station for decision support by detecting and classifying the heartbeat. Heartbeat detection in an ECG primarily depends on the QRS complex. However, QRS detection by itself is not sufficient for heartbeat detection, which must precede the recognition of features for detecting arrhythmia. One of the significant contributions of this study is the method of detecting the P-wave using the QRS complex, which is in turn detected by the Pan-Tompkins algorithm [7] (consisting of a band-pass filter, a differentiator and an integrator over a moving window). Eventually, the use of the P-wave with the QRS complex leads to the accurate detection of heartbeats related to arrhythmia.

3.1.4. Monitoring Station

The monitoring station (MS) provides a graphical user interface. Therefore, both the patient and the expert operators can remotely check the signal. The patient, experts and a few others who have access rights can view the ECG, heart rate and patient profile through the web or a mobile device application. Moreover, experts using the MS can also view and determine the signal conditions of many patients through the web application. If an expert identifies a dangerous condition, then the MS immediately contacts medical services.

3.2. Threat Model

In the scope of this work, we consider normal users, including servers, as potential adversaries. Adversaries have limited capabilities, in that they can only access publicly-available information, including information from wireless communications. This is no more information than can be accessed by normal users of the ECG monitoring system. Thus, we consider that adversaries do not have the ability to distinguish the originator of data eavesdropped from a wireless communication. In the case of servers, we assume an honest-but-curious model that accurately follows a provided protocol, but may attempt to obtain information from the communication session or stored data. In addition, we do not consider that the adversaries can easily compromise other entities in order to obtain the identity of users or other valuable information. Based on these conditions, the following attacks can be carried out by adversaries.

- **Eavesdropping:** Essentially, we assume that an attacker has the ability of eavesdropping on data during transmissions. Wireless communications, including Bluetooth, which is used to transmit the sensing data, are vulnerable, because the transmitted data can be leaked by the major function hooking scheme of the windows kernel driver [26,27]. Thus, an attacker close to the sensing system can eavesdrop on the sensing data and the identity of a user via the weak point of our system.
- **Leaking:** We treat the DSS and service provider as potential attackers. The ECG analysis results and identity of an individual user could be leaked by the service providers or system managers of the DSS and MS, respectively. We only focus on cases where health information is leaked along with the identity of the relevant user.
- **Tracking:** After obtaining data (raw ECG data or an ECG analysis) by eavesdropping or leaking, an attacker could attempt to determine the relation between the data and the user. In addition, the attacker could attempt to trace a particular user by determining the relations between pseudonyms.

The collected sensing data for detecting arrhythmias is personal information for the user, and it could be used by an attacker in various ways. One possible scenario is selling the collected data for profit. Pharmaceutical companies or clinics could use such data for targeted marketing. In addition, the patient may not want his/her illness to be known to others. Such types of attacks are only possible if the attacker could obtain both the health information and the identity of the user.

3.3. Security Objectives

- Depending on the system and threat models described earlier, we define two security objectives.
- Communication security: An attacker close to a user cannot obtain sensing signals from the wireless communication between the MCU and the smart device, nor between the smart device and the DSS.
 - Privacy preservation: An attacker cannot establish a link between a particular user and their sensing signal or analysis results from the DSS. In the case of an emergency, the personal information of a user can be disclosed. During this step, an attacker can obtain only a limited amount of data from the server. In other words, from a collection of personal information provided and a pseudonym, the attacker cannot distinguish the ECG information of a particular user.

A simple method of preserving user privacy is to employ a pseudonym as an anonymous identity for communication, because this can easily hide the relation between data and a user's identity. However, an attacker can easily obtain the identity of a certain user through long-term observation. Many previous schemes have mentioned the traceability of a single pseudonym [28–31]. In our system model, the attacker can obtain contact information for the pseudonym in an emergency. All of the patient's information from before and after the emergency will be leaked if a single pseudonym-based approach is applied, and thus, the security model has to deal with forward and backward privacy. To deal with an emergency, a user's identity should be revealed to receive medical service. Otherwise, an inside attacker can determine the relation between the user's identity and the sensing signal. The forward privacy indicates that the attacker cannot determine a relation between sensing signals after an emergency. The backward privacy indicates that the attacker cannot determine a relation between sensing signals before the emergency.

4. System Architecture for Privacy-Preserving Intelligent ECG Monitoring

In this section, we propose an intelligent ECG monitoring system incorporating privacy preservation. The proposed scheme comprises two stages. First, we design a secure sensing signal encryption scheme using a conventional public key cryptosystem, to protect the sensing data transferred from the sensor to the MS via a smart device. Second, we present an AID scheme to preserve the user privacy during the arrhythmia recognition process. This also hides the user information from experts at the MS, but this can be revealed in an emergency. Before describing the proposed scheme, we present the notations used in this paper in Table 1.

Table 1. Notations.

Notation	Description
k	A symmetric session key
(pk, sk)	Public, private key pair
p_ℓ	A pseudonym
PS	A set of pseudonyms, $PS = \{p_1, p_2, \dots, p_\ell\}$
a_ℓ	An anonymous ID
A_{ID}	A set of anonymous IDs, $A_{ID} = \{a_j\}_{1 \leq j \leq \ell}$
$H(\cdot)$	A cryptographic hash function
$PRG(\cdot)$	A pseudo-random number generator
$E_k \{\}$	An encryption function using key k
$S_k \{\}$	A signature function using key k
KS	A key stream
ES	An encrypted ECG signal using key k

4.1. Setup

Each user generates and uses a key and a pseudonym for each session to ensure privacy. One session is the duration of a pseudonym and the corresponding key. By changing these frequently,

the system can provide stronger privacy protection. The length of each session is flexible. If a user wants a higher level of privacy, he/she can adjust to make the duration shorter. The user also can adjust to make the duration longer in the case that they require efficient operation.

When a user first establishes a connection between the MCU of a body sensor network and a smart device, the MCU and the smart device share a symmetric session key $\{k_s\}$ to deal with the limited storage of the MCU.

The RA, DSS, MS and the smart device of a user use a public key cryptosystem to establish secure communications. The public/private key pairs for the entities are (pk_{RA}, sk_{RA}) , (pk_{DSS}, sk_{DSS}) , (pk_{MS}, sk_{MS}) and (pk_U, sk_U) , respectively.

4.2. Revocation Authority

Each user has a unique ID , and the user registers this along with his/her contact information to deal with emergencies. To generate an AID, the user encrypts their ID using a public key of the RA and sends this to the RA. At this point, we employ a set of session pseudonyms such that each user uses one pseudonym for only one session, to deal with forward and backward privacy in the single pseudonym approach [31].

The RA generates a set of unique pseudonyms $PS = \{p_1, p_2, \dots, p_\ell\}$, and computes the AID $AID = \{a_j\}_{1 \leq j \leq \ell}$ with its signature as:

$$a_j = E_{pk_U}\{H(ID||p_j)\}, S_{sk_{RA}}\{a_j\}, 1 \leq j \leq \ell$$

where ℓ is a natural number greater than one, which is defined by the user. To address the trade-off between privacy and efficiency, where the privacy is enhanced by changing the pseudonyms more frequently, we allow the user to define the frequency and ℓ while using the application. However, each session should be longer than 30 minutes to ensure an accurate analysis result.

The RA stores (ID, AID) pairs, encrypts these using the user's public key and then sends them to the user's smart device. At this point, our system only uses the public key to encrypt the (ID, AID) pairs. Therefore, an attacker cannot obtain the AID used for an actual communication between the user and the DSS. Thus, using the user's public key does not affect the privacy of the system. Finally, the user sends sensing signals to the DSS, preserving privacy by using one of the AIDs. As the user cannot receive services when an incorrect AID is used, a legal AID that is generated by the RA should be entered. After receiving the AID , the user generates a set of keys $\{k_i\}_{a \leq i \leq \ell}$.

In case of an emergency, the MS sends the AID to the RA. The RA then determines the patient information from the AID list and performs the step of connecting the patient with a doctor. Our scheme employs the RA as an additional entity to decentralize secret information and increase the effectiveness of privacy preservation.

4.3. ECG Sensing

To transfer the sensing signal securely, we use a stream cipher that consists of a pseudo-random generator (PRG) and a bitwise exclusive OR operator [32]. A sensor installed on the human body typically consists of a resource-constrained and battery-powered device. Moreover, considerable computations result in the production of more heat, which can be problematic in terms of patient safety. Furthermore, a body sensor requires a powerful processor and a large capacity battery to apply cryptography to its various functions, which would increase the size of the body sensing system and interfere with its day-to-day usage. Accordingly, we propose a simple and secure method that minimizes the encryption overhead and protects sensing data during transfer.

The MCU of the user generates the keystream KS_s using the PRG such that $KS_s = PRG(k_s)$. Then, it computes the exclusive OR operation for the ECG signal SS_i as $ES_s = KS_s \oplus SS_i$, where ES_s denotes the encrypted ECG signal. After receiving this, the smart device of the user performs the same

operation as the MCU to decrypt the ECG signal, and the user can monitor their own ECG on selected smart devices.

To send the sensing signal safely, the user generates a key stream KS_i using the PRG such that $KS_i = PRG(k_i)$ and encrypts SS_i using KS_i as $ES_i = KS_i \oplus SS_i$. Subsequently, the user selects an a_i randomly and sends the received signal stream to the DSS to detect arrhythmias using the ECG. The first time the user employs a choice of a_i , they send $a_i, S_{sk_{RA}}\{a_i\}$ and $E_{pk_{DSS}}\{k_i\}$ for validity. Subsequently, the user sends the signal stream to the DSS as $a_i||ES_i$.

For privacy reasons, we use a set of pseudonyms and a set of keys. If a user only uses a single pseudonym or key, then the DSS can easily trace that user. When a user's identity is revealed in an emergency, the DSS can obtain all of the stored ECG signals and analysis results, as well as future information for the user. Therefore, users should periodically change their pseudonyms and keys to minimize leaked data after an emergency. After a received pseudonym is exhausted, the user requests a novel set of pseudonyms from the RA.

The DSS detects arrhythmias using continuous ECG streaming from users, and the pseudonyms transform that stream into a discontinuous signal. Thus, we should consider the effect of pseudonyms on the detection process. Consequently, we describe the simulation results of our scheme in Section 6, based on 30-minute of ECG signals, and show that we can ensure 96.63% accuracy if the pseudonym changing period is longer than half an hour.

4.4. Analysis of an ECG for Arrhythmia Detection

The DSS should first decrypt the signal stream to analyze the ECG. To achieve this, the DSS decrypts the key signal using its private key and obtains k_i . By obtaining k_i , the DSS can generate the same KS_i as the MCU and decrypt the ECG signal as $SS_i = KS_i \oplus ES_i$. Using SS_i , the DSS analyzes the multiple aspects of the ECG.

4.4.1. Heartbeat Detection and Feature Extraction

First, heartbeats are detected using the determined QRS complexes and P-waves. Figure 4 depicts the process of the Pan–Tompkins algorithm, and Figure 5 [25] illustrates the results of feature extraction, with the step-by-step output of the algorithm performed on Record 200 in the MIT-BIH arrhythmia database. Figure 5a shows the original ECG signal. The original signal is normalized by the mean value, as shown in Figure 5b. The band-pass filter is created by combining a low-pass filter with a high-pass filter. This reduces noise such as muscle noise, 60-Hz interference, baseline wander and T-wave interference in the ECG signal. The differential equation for the low-pass filter is:

$$\begin{aligned} y(nT_s) &= 2y(nT_s - T_s) - y(nT_s - 2T_s) + x(nT_s) \\ &\quad - x(nT_s - 6T_s) + x(nT_s - 12T_s), \end{aligned} \quad (1)$$

where T_s denotes the sampling period, x is the amplitude of the n -th ECG sample and y is the amplitude after filtering. The difference equation for the high-pass filter is:

$$\begin{aligned} y(nT_s) &= 32x(nT_s - 16T_s) \\ &\quad - [y(nT_s - T_s) + x(nT_s) - x(nT_s - 32T_s)]. \end{aligned} \quad (2)$$

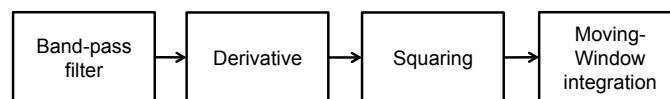


Figure 4. Process of the Pan–Tompkins algorithm.

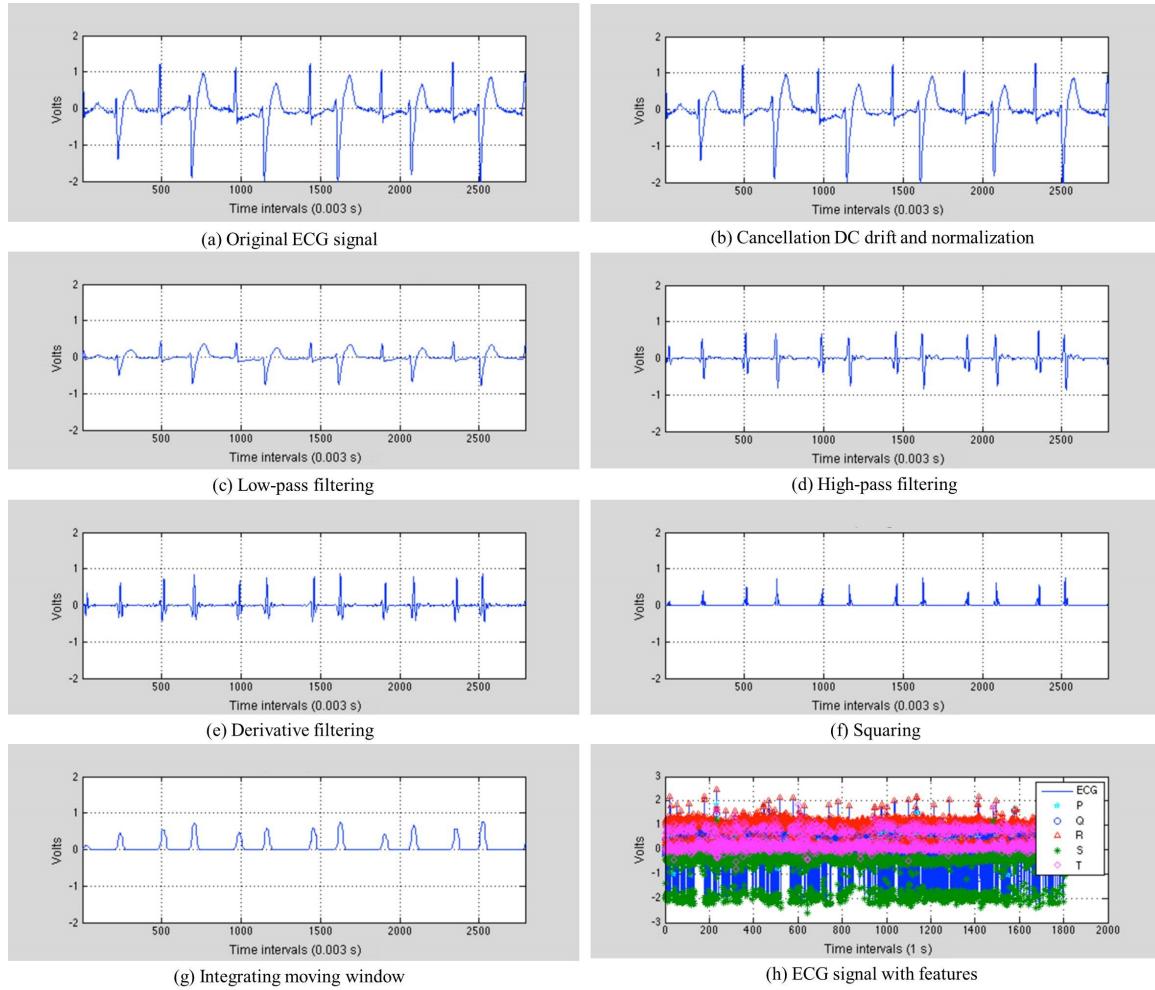


Figure 5. Results of feature extraction with step-by-step output of the Pan–Tompkins algorithm.

The DSS sets the low-pass filter with a cutoff frequency of 11 Hz and the high-pass filter with a cutoff frequency of 5 Hz, as shown in Figure 5c,d. After being filtered, the ECG signal is differentiated to provide slope information using the following differential equation:

$$\begin{aligned} y(nT_s) = & \frac{1}{8T_s} [-x(nT_s - 2T_s) - 2x(nT_s - T_s) \\ & + 2x(nT_s + T_s) + x(nT_s + 2T_s)], \end{aligned} \quad (3)$$

Equation (3) approximates the ideal derivative of frequencies up to 30 Hz, and Figure 5e presents the results of the derivative. This is then squared point by point, making all of the data points in the processed signal positive and emphasizing the higher frequencies, as shown in Figure 5f. The differential equation for this squaring is:

$$y(nT_s) = [x(nT_s)]^2. \quad (4)$$

Integrating the moving window provides waveform feature information, which is added to the slope of the R-wave. This is achieved using the following differential equation:

$$\begin{aligned} y(nT_s) = & \frac{1}{N_s} [x(nT_s - (N_s - 1)T_s) \\ & + x(nT_s - (N_s - 2)T_s) + \dots + x(nT_s)], \end{aligned} \quad (5)$$

where N_s is the number of samples. This produces “re-echo mountaintops”, as depicted in Figure 5g, where each of the peaks corresponds to a heartbeat. The DSS finds the QRS complex in the heartbeat, where Q is the starting point, S is the ending point and R is the peak. The DSS then determines the P-wave using the QRS complex. This is observed between the S point of the current heartbeat and the Q point of the next heartbeat. The DSS divides this range into two, and uses the peak in the second subrange as the point P. Figure 5h depicts the sample results for the QRS complex and P-wave detection from Record 200 in the MIT-BIH arrhythmia database. The QRS complex and P-waves observed are all used for extracting the features from the heartbeats, and descriptions of these features are provided in Table 2.

Table 2. Description of the features used for heartbeat classification.

Feature	Description
P-position	Position of P point
Q-position	Position of Q point
R-position	Position of R point
S-position	Position of S point
P-value	Value of P point (amplitude)
Q-value	Value of Q point (amplitude)
R-value	Value of R point (amplitude)
S-value	Value of S point (amplitude)
PR-distance	Distance between P point and R point
RR-interval	Distance between consecutive R-waves
MHR	Maximum heart rate
BHR	Resting heart rate

The RR interval is used for assessing the ventricular rate. The DSS calculates the heart rate using the RR interval value. This is typically calculated using one of two algorithms, either the Fox and Haskell formula [33] or the mathematical formula described in [34]. The heart rate provides an easy cardiovascular measurement, particularly in comparison with invasive and noninvasive procedures used to estimate the stroke volume and cardiac output. To provide accurate health information, we display the maximum and resting heart rates. The maximum heart rate is the highest heart rate achieved during maximal exercise and is calculated as follows:

$$MHR = 220 - age. \quad (6)$$

There are many methods for calculating the predicted maximum heart rate, including those of Tanaka et al. Fox and Haskell, Robergs and Landwehr, Gulati and Lund. We use Fox and Haskell's formula, because this is widely adopted for the heart rate (max). The basal or resting heart rate is measured as follows:

$$BHR = 60 / RRinterval, \quad (7)$$

when the patients are relaxed, but awake in a naturally temperate environment, and should have neither recently exerted themselves nor been subject to stress or a surprise. The DSS applies this basal heart rate formula.

4.4.2. Learning Different Types of Heartbeats Using a C4.5 Algorithm

The DSS classifies the individual heartbeats in the ECG based on decision tree learning, which is one of the most widely-employed classification techniques. Its classification accuracy is competitive with other learning methods, and it is considerably efficient. The learned classification model is represented as a tree, called a decision tree. We trained the decision tree based on the C4.5 algorithm from Quinlan [8], which can provide prominent results, readability, flexibility and efficiency.

The Iterative Dichotomiser 3 (ID3) [35] algorithm is mostly used for training the decision tree. This makes statistical-based decisions and is therefore less sensitive to errors in individual training examples. Although the ID3 algorithm is used in various domains, it is not applied to our server, because the discrete-valued target function domain cannot be applied to the ID3 algorithm. Therefore, our server uses the C4.5 algorithm, which is an extension of the basic ID3 algorithm. The pseudocode of this learning algorithm is presented in Algorithm 1.

Algorithm 1 C4.5 algorithm for arrhythmia learning.

```

1: function DecisionTree( $D, A, T$ )            $\triangleright D$ : training dataset,  $A$ : feature set,  $T$ : decision tree
2:   if  $D$  contains only training examples of the same heartbeat type  $c_j \in C$  then
3:     Make  $T$  a leaf node labeled with heartbeat type  $c_j$ ;                       $\triangleright C$ : heartbeat types
4:   else if  $A = \emptyset$  then
5:     Make  $T$  a leaf node labeled with heartbeat type  $c_j$ , which is the most frequent heartbeat type
      in  $D$ ;
6:   else                                 $\triangleright D$  contains examples belonging to a mixture of heartbeat types
7:      $p_i=\text{impurityEval-1}(D)$ ;  $\triangleright$  We select a single feature to partition  $D$  into subsets so that each
      subset purer
8:     for Each feature  $A_i \in A (= \{A_1, A_2, \dots, A_k\})$  do
9:        $p_i=\text{impurityEval-2}(A_i, D)$ ;
10:      end for
11:      Select  $A_g \in \{A_1, A_2, \dots, A_k\}$  that provides the biggest impurity reduction, computed using
       $p_0 - p_i$ ;
12:      if  $(p_0 - p_g) < \text{threshold}$  then           $\triangleright A_g$  does not significantly reduce impurity  $p_0$ 
13:        Make  $T$  a leaf node labeled with  $c_j$ , the most frequent heartbeat type in  $D$ ;
14:      else                                 $\triangleright A_g$  is able to reduce impurity  $p_0$ 
15:        Make  $T$  a decision node on  $A_g$ ;
16:        Let the possible values of  $A_g$  be  $v_1, v_2, \dots, v_m$ . Partition  $D$  into  $m$  disjoint subsets
       $D_1, D_2, \dots, D_m$  based on the  $m$  values of  $A_g$ ;
17:        for Each  $D_j \in D (= \{D_1, D_2, \dots, D_m\})$  do
18:          if  $D_j \neq \emptyset$  then
19:            Create a branch (edge) node  $T_j$  for  $v_j$  as a child node of  $T$ ;
20:            DecisionTree( $D_j, A - \{A_g\}, T_j$ );
21:          end if
22:        end for
23:      end if
24:    end if
25:  end function
  
```

In particular, a decision tree T simply partitions the training dataset D into disjoint subsets, so that each subset is as pure as possible (of the same heartbeat type), by considering a feature set A . The learning of a tree is typically performed using the divide-and-conquer strategy, which recursively partitions the data to produce the tree. Initially, all of the examples are at the root. As the tree grows, the examples are subdivided recursively. In this study, we assume that every feature in D takes discrete values.

The stopping criteria of the recursion are presented in Lines 1–4 in Algorithm 1. The algorithm stops when all of the training examples in the current data are of the same heartbeat type, or when every feature is used along the current tree path. In tree learning, each successive recursion selects the best feature to partition the data at the current node according to the values of the feature. The best feature is selected based on a function that aims to minimize the impurity after the partitioning (Lines 7–11). In other words, it maximizes the purity. The key in decision tree learning is thus the choice of the impurity function, which is employed in Lines 7, 9 and 11 in Algorithm 1. The recursive recall of the algorithm is given in Line 20, which takes the subset of the training examples at

the node for further partitioning to extend the tree. This is a greedy algorithm, with no backtracking. The process of node creation is irreversible; no modification is possible once a node is created.

The most popular impurity functions used for decision tree learning are information gain and information gain ratio, which are used in C4.5 as two portions. Let us first discuss information gain, which can be extended slightly to produce information gain ratio. The information gain measure is based on the entropy function from information theory:

$$\text{Entropy}(D) = - \sum_{j=1}^{|C|} \Pr(c_j) \log_2 \Pr(c_j), \text{ where } \left(\sum_{j=1}^{|C|} \Pr(c_j) = 1 \right). \quad (8)$$

$C = \{c_1, c_2, \dots, c_{|C|}\}$ is a set of $|C|$ heartbeat types and $\Pr(c_j)$ is the probability of heartbeat type c_j to occur in the dataset D . This is given by the number of examples of heartbeat type c_j in D divided by the total number of examples in D .

4.4.3. Heartbeat Classification for Arrhythmia Detection Using a Decision Tree

The DSS classifies the individual heartbeats in the ECG using a decision tree learned by the C4.5 algorithm, as mentioned above, and then sends the AID and analysis results to the MS. Each of the ECG records from the MIT-BIH arrhythmia database can contain a maximum of 11 heartbeat types, consisting of a normal beat (N) and 10 abnormal beats: left bundle branch block beat (L), right bundle branch block beat (R), atrial premature beat (A), aberrated atrial premature beat (a), premature ventricular contraction beat (V), fusion of ventricular and normal beat (F), ventricular flutter wave beat (!), atrial escape beat (e), ventricular escape beat (E) and paced beat (P). Table 3 summarizes the 10 different types of ventricular and atrial heartbeat that we are targeting to classify for arrhythmia detection, which are all annotated for signal quality and rhythm changes.

Table 3. Description of heartbeat types.

Arrhythmia		The Rest	
Type	Description	Type	Description
L	Left bundle branch block beat	N	Normal beat
R	Right bundle branch block beat	?	Beat not classified during learning
A	Atrial premature beat	*	Change in signal quality
a	Aberrated atrial premature beat		Isolated QRS-like artifact
V	Premature ventricular contraction beat	+	Rhythm change
F	Fusion of ventricular and normal beat		
!	Ventricular flutter wave beat		
e	Atrial escape beat		
E	Ventricular escape beat		
P	Paced beat		

5. Implementation

5.1. Mobile System

We designed an actual system for the simulation. As our scheme includes a real-time streaming signal with a feature extraction and classification process performed by a server, it is important to prove that our scheme can operate in real applications. Therefore, our implementation focused on two concerns. First, as mentioned earlier, we implemented our scheme to show that it is applicable to the remote monitoring setting. Second, our implementation shows that it can detect and classify heartbeats with a high accuracy.

The system software consisted of a mobile system and a DSS. The mobile system was implemented in Java, using the Android SDK 2.3.3 (Google, Mountain View, CA, USA). We selected the Android

platform because it is open source. Moreover, it has a large developer community that writes applications for it and provides code portability. We selected Apache 2.0 (Apache Software Foundation, Wakefield, MA, USA) to host the web application, which links to MySQL. Figure 6 illustrates the implementation of the mobile system.



Figure 6. Results of mobile device implementation.

5.2. DSS Specification

The DSS connects with an individual database and analyzes the heartbeat based on the QRS complexity algorithm. It displays graphical ECG waveforms through a web application. Figure 7 depicts the results from the web application. We employed several frameworks for the implementation of the DSS, including PHP 5.1.4 (Zend Technologies Ltd. Louisville, CO, USA), pChart 2.1.3 and MySQL 5.0.51a (Oracle Corporation, Santa Clara, CA, USA). We used PHP as the primary framework, owing to its operational speed and cross-platform compatibility. As a PHP class, pChart is an oriented framework designed for creating aliased charts. We used a MySQL database to maintain records for the patients and experts affiliated with the health monitoring program.

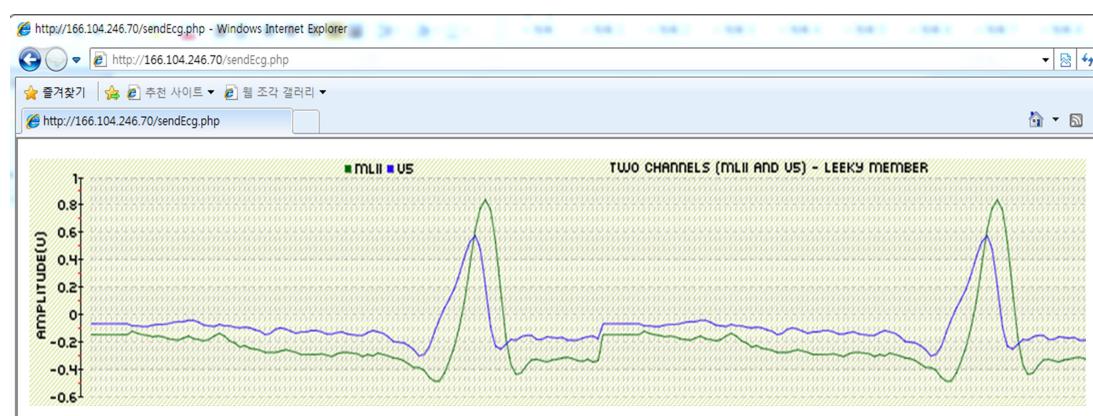


Figure 7. Results of web application implementation.

6. Evaluation and Results

In this section, we evaluate our scheme based on the three criteria of security, efficiency and accuracy.

6.1. Security

Concerning the security and privacy of the ECG monitoring system, we adopted a PRG-based stream cipher to protect the ECG signal, and we conceal the identity of the user from the DSS using a pseudonym system. To verify that our system is secure, we verify that our scheme satisfies the security objectives.

We define the indistinguishability of encryptions from the perspective of communication security, which is the first security objective.

Definition 1 (Indistinguishability of encryptions [36]). *An encryption scheme, (G, E, D) , which consists of a key generator G , encryption E and decryption D , exhibits indistinguishable encryptions. For every polynomial-size circuit family $\{C_n\}$, every positive polynomial p , all sufficiently large n and every $x, y \in \{0, 1\}^{poly(n)}$ (i.e., $|x| = |y|$),*

$$|Pr[C_n(E_{G(1^n)}(x)) = 1] - Pr[C_n(E_{G(1^n)}(y)) = 1]| < \frac{1}{p(n)}.$$

The probability in these terms is taken over the internal coin tosses of the algorithms G and E .

A symmetric key encryption is semantically secure if and only if it exhibits indistinguishable encryption [36]. Thus, we prove that our encryption scheme has indistinguishability.

Claim 1. Suppose that there exist PRGs that exhibit robustness against polynomial-size circuits, and our scheme adopts one of these. Then, the encryption scheme based on the PRG satisfies Definition 1.

Proof. First, we formally explain our encryption scheme. The key for the security parameter n is a uniformly-distributed n bit-long string, denoted by k_i . To encrypt a sensing stream SS_i , the encryption algorithm uses the key k_i as a seed for a PRG, denoted by g , which stretches seeds of length n into sequences of length $|SS_i|$. The ciphertext is obtained by a bit-by-bit exclusive OR of SS_i and $g(k_i)$. Decryption is performed in the same manner.

We show that this encryption scheme satisfies Definition 1. Intuitively, this follows from the hypothesis that g is a PRG and the fact that $SS_i \oplus g(U_n)$ is uniformly distributed over $\{0, 1\}^{|SS_i|}$. In particular, in order to obtain a contradiction, suppose that for some polynomial-size circuit family $\{C_n\}$, a polynomial p and infinitely many n 's, it holds that:

$$|Pr[C_n(SS_i \oplus g(U_n)) = 1] - Pr[C_n(SS_j \oplus g(U_n)) = 1]| > \frac{1}{p(n)},$$

where U_n is uniformly distributed over $\{0, 1\}^n$ and $|SS_i| = |SS_j| = m = poly(n)$. On the other hand,

$$Pr[C_n(SS_i \oplus g(U_n)) = 1] = Pr[C_n(SS_j \oplus g(U_n)) = 1].$$

Thus, without loss of generality:

$$|Pr[C_n(SS_i \oplus g(U_n)) = 1] - Pr[C_n(SS_i \oplus U_m) = 1]| > \frac{1}{2 \cdot p(n)}.$$

By incorporating SS_i into the circuit C_n , we obtain a circuit that distinguishes U_m from $g(U_n)$, in contradiction of our hypothesis regarding the pseudorandomness of g . \square

Therefore, the stream cipher derived from the PRG is semantically secure [36]. The only problem regarding the security of our system is the key. If the key stored in the MCU and smart device is leaked, then a third party could obtain the ECG signal and the relationship between the signal and the user. To store the key securely, tamper-resistant memory can be attached to the MCU. In addition, key leakage resulting from weaknesses in the smart device system, such as malware infections, constitutes a security issue for the device itself. The treatment of this is beyond the scope of this study.

Next, we verify that our scheme can preserve user privacy. This privacy preservation is the second security objective.

Definition 2 (Polynomial-time indistinguishability of pseudonym [37]). *Suppose that there exist PRGs that exhibit robustness against polynomial-size circuits. Then, a pseudonym generated using such a PRG exhibits polynomial-time indistinguishability.*

Two pseudonyms $X \stackrel{\text{def}}{=} \{X_n\}_{n \in \mathbb{N}}$ and $Y \stackrel{\text{def}}{=} \{Y_n\}_{n \in \mathbb{N}}$ that are uniformly distributed over $\{0, 1\}^n$ by the PRG are indistinguishable in polynomial time if for every probabilistic polynomial-time algorithm D , every positive polynomial $p()$ and all sufficiently large n 's, it holds that:

$$|Pr[D(X_n, 1^n) = 1] - Pr[D(Y_n, 1^n) = 1]| < \frac{1}{p(n)}.$$

Our system only provides the information required to detect arrhythmia, such as age. The AIDs used for privacy preservation are generated by the PRG, and thus, the AIDs exhibit polynomial-time indistinguishability according to Definition 2. Therefore, our scheme is secure against tracking attacks, as well as data leakage resulting from the DSS. However, the DSS and MS will be able to obtain a user's identity from the RA when an instance of arrhythmia is detected. In this case, they can obtain the real identity of a user and their ECG signals that are stored in the server. In terms of privacy, changing pseudonyms is useful for minimizing the data leakage, because only the ECG data stored under the current pseudonym will be leaked in an emergency. If an expert requires further information regarding the user, the expert can request additional pseudonyms from the RA. In addition, another security issue is the manipulation of ciphertexts. An attacker could manipulate the service by transferring ciphertexts in order to disrupt a smooth use of services. This problem can occur not only under the XORing circumstance, but also under other encryption circumstances, as this is a problem of the networks. Therefore, we do not consider this problem in this paper. Such integrity issues could easily be addressed by periodically applying a message authentication code (MAC).

6.2. Efficiency

To evaluate the efficiency of our scheme, we simulated it against other well-known encryption schemes. We implemented our experiment on an Intel (R) core (TM) 2 Quad processor, running at 2.66 GHz, 4.00 GB of RAM and an HDD Serial ATA 3.0 Gbit/s drive with an 8 MB buffer. All algorithms are implemented using Python 2.7.9 on Linux Ubuntu 12.04 LTS 32 bit (Canonical Ltd., London, UK), and we used the PyCrypto library version 2.6.1 for the ciphers (AES, DES and RC4). We used a 128-bit key with cipher feedback (CFB) mode for AES and a 64-bit key with CFB mode for DES. For the PRG, we used “os.urandom()”, which is the basic library of Python that is suitable for cryptographic usage. We measured the encryption and decryption time for 300 s of an ECG signal. The volume of this ECG signal is approximately 4 Mbyte. However, this could vary in accordance with the attached sensors. All experimental results represent the average over 10 trials.

Table 4 shows the comparison of the encryption and decryption times between ciphers. The results show that our scheme is considerably efficient compared with most of the other cryptography schemes, although RC4, which consists of a key scheduling algorithm and bit-wise exclusive OR, exhibits a similar efficiency to our scheme. However, RC4 is known to be insecure, owing to a weakness of the key scheduling algorithm [38]. Therefore, it is better to use a cryptographic PRG as a key

expanding algorithm. In this case, a developer can select a PRG algorithm based on the condition of an MCU.

Table 4. Simulation results (ms).

	AES	DES	RC4	Ours
Encryption	1220.54	1366.30	47.09	46.80
Decryption	1222.32	1362.98	48.86	45.64

In addition, we only used RSA with a 1024-bit key to encrypt the set of pseudonyms for mobile devices, not for body sensors. This requires 1.416 ms of computational overhead for encryption and 17.97 ms for decryption. As the pseudonym communication is rarely used, we can say that this overhead is reasonable for privacy preservation.

6.3. Accuracy of Arrhythmia Classification

6.3.1. ECG data

To evaluate our classifier in terms of accuracy, we used the MIT-BIH arrhythmia database [14], which contains 48 30-minute excerpts of two-channel ambulatory ECG recordings, obtained from 47 patients studied at the BIH Arrhythmia Laboratory between 1975 and 1979. Twenty-three recordings were selected at random from a set of 4000 24-hour ambulatory ECG recordings, collected from a mixed population of inpatients (approximately 60%) and outpatients (approximately 40%) at Boston's Beth Israel Hospital. The remaining 25 recordings were selected from the same set, to include less common, but clinically-significant arrhythmias (such as complex ventricular, junctional and supraventricular arrhythmias and conduction abnormalities) that would not be represented well in a small random sample.

The recordings were digitized at 360 samples per second per channel, with 11 bit resolution, over a 10-mV range. Two or more cardiologists independently annotated each record, and disagreements were resolved to obtain computer-readable reference annotations for each beat (approximately 110,000 annotations in all) included in the database. We considered nine records from the database, numbered as 106, 107, 109, 115, 122, 207, 212, 221 and 233. Table 5 provides a statistical overview of these heartbeat types for the nine records. The column headed @ contains totals for the following events: beat not classified during learning (?), change in signal quality (*), isolated QRS-like artifact (!) and rhythm change (+).

Table 5. Characteristics of each record selected from the MIT-BIH arrhythmia database.

Record Number	N	L	R	A	a	V	F	!	e	E	P	@
106	1507	-	-	-	-	520	-	-	-	-	-	71
107	-	-	-	-	-	59	-	-	-	-	2078	3
109	-	2492	-	-	-	38	2	-	-	-	-	3
115	1953	-	-	-	-	-	-	-	-	-	-	9
122	2476	-	-	-	-	-	-	-	-	-	-	3
207	-	1457	86	107	-	105	-	472	-	105	-	53
212	923	-	1825	-	-	-	-	-	-	-	-	15
221	2031	-	-	-	-	396	-	-	-	-	-	35
223	2029	-	-	72	1	473	14	-	16	-	-	38

6.3.2. Evaluation

We used the three standard metrics of sensitivity (*Se.*), specificity (*Sp.*) and accuracy (*Acc.*) to quantify the server performance. Sensitivity refers to the ability of a test to correctly identify the results of a classified heartbeat with a positive result and is defined as follows:

$$Se. = N_{TP} / (N_{TP} + N_{FN}), \quad (9)$$

where N_{TP} represents the number of true positives and N_{FN} represents the number of false negatives. Specificity refers to the ability of a test to correctly identify the results of a classified type without a positive result and is given by:

$$Sp. = N_{TN} / (N_{TN} + N_{FP}), \quad (10)$$

where N_{TN} represents the number of true negatives and N_{FP} represents the number of false positives. Accuracy refers to the ability of a test to correctly identify the results of a classified type both with and without positive results. It reflects both the sensitivity and specificity and is expressed as:

$$Acc. = (N_{TP} + N_{TN}) / (N_{TP} + N_{TN} + N_{FN} + N_{FP}). \quad (11)$$

We evaluated the performance of the server in two aspects for the analysis of the heartbeat. The first aspect of the evaluation related to heartbeat detection, which affects the classification performance. The 11 types of heartbeat listed in Table 3 were considered for evaluation, and Table 6 illustrates the performance for each of the nine records. For the nine records we considered from the MIT-BIH arrhythmia database, the server achieved an overall accuracy of 95.74%, with a sensitivity of 97.21% and a specificity of 94.26%. It is noted that Record 107 exhibits many abnormal beats, with an unusually large, typically peaked P-wave, which resulted in a particularly low rate of specificity of 50.05%. This result could be enhanced by a deeper analysis of the P-wave. However, this is beyond the scope of the present study.

The second aspect of the evaluation concerned heartbeat classification for arrhythmia detection. We used a 10-fold cross-validation method to test the classification [39]. The classification performance is summarized in Table 6. The server achieved an overall accuracy of 96.63%, with a sensitivity of 95.44% and a specificity of 97.81%. If we consider an ECG recorded for a longer period of time, the analysis for the diagnosis could be more accurate.

Table 6. Results of heartbeat detection and classification for arrhythmia recognition (sensitivity (*Se.*), specificity (*Sp.*) and accuracy (*Acc.*)).

Rec.	Heartbeat Detection			Arrhythmia Detection		
	Se. (%)	Sp. (%)	Acc. (%)	Se. (%)	Sp. (%)	Acc. (%)
106	93.99	99.80	96.90	95.80	98.00	96.90
107	99.86	50.05	74.96	93.00	98.30	95.65
109	99.64	99.84	99.74	93.30	94.40	93.85
115	99.59	100.00	99.80	97.80	100.00	98.90
122	99.96	100.00	99.98	99.90	100.00	99.95
207	86.83	99.90	93.37	92.60	98.80	95.70
212	99.53	99.85	99.69	94.20	97.80	96.00
221	98.21	99.67	98.94	94.20	95.70	94.95
233	97.30	99.22	98.26	98.20	97.30	97.75
Average	97.21	94.26	95.74	95.44	97.81	96.63

7. Conclusions

In this paper, we have proposed a privacy-preserving intelligent ECG monitoring system for early arrhythmia detection and described its implementation. The major steps required for accurate

recognition of arrhythmia are (1) the accurate detection of heartbeats and (2) defining the significant features of those heartbeats and extracting them to recognize various types of heartbeats, which involves leveraging the effectiveness of a machine learning algorithm and employing it for decision making. The steps we invoked to achieve the required accuracy are as follows. First, we proposed and applied a scheme for P-wave detection, along with the well-known Pan–Tompkins algorithm, to enhance the accuracy of heartbeat detection at the first step. Second, this enhanced accuracy in the heartbeat detection naturally has a positive influence on the subsequent feature extraction process. Third, we introduced a classifier based on a decision tree and trained it using a C4.5 algorithm, which resulted in several enhancements to the former ID3 algorithm, by considering various combinations of ECG features. We rigorously investigated the effects of these combinations on the accuracy of the decision making using a decision tree and determined one of the feature sets featuring the best compromises by using a brute force search in the third step. As a result, we achieved an average accuracy of 96.63% in recognizing arrhythmia for nine example records from the MIT-BIH arrhythmia database that we considered for our experiments, which is superior to the results (95.8%) achieved in our preliminary studies.

In addition, we designed a simple and effective sensing data scrambling scheme to ensure the security of streamed sensing data and adopted AIDs to preserve user privacy. Using our proposed system, medical experts can capture the intermittent ECG waveforms that may reveal or lead to a more serious problem. An accurate diagnosis of arrhythmia based on a heartbeat detection and classification algorithm can be achieved in real time, which is beneficial for remote medical care. This system represents a low-cost solution, which could be affordable across medical environments.

It is clear that no scheme of this sort is fully proven until it has been tried in practice. However, the implementation and testing of medical systems is especially fraught with difficulties, from patient confidentiality to legal concerns, and cannot reasonably be undertaken until there is as much confidence as possible in the techniques and associated parameters being deployed. It is this confidence that we aim to build in this paper, through simulations based on offline data.

Acknowledgments: This research was partially supported by the Ministry of Science, ICT and Future Planning (MSIP), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-2014-0-00636) supervised by the IITP (Institute for Information & Communications Technology Promotion). This research was also supported in part by the IITP grant funded by the Korean government (MSIP) (2014-0-00065, Resilient Cyber-Physical Systems Research; 2017-0-00395, Information management system for dark web scanning). This research was also supported in part by the NRF (National Research Foundation of Korea) grant funded by the Korean government (MEST)(NRF-2015R1D1A1A09058200).

Author Contributions: J.S. and K.K. conceived and designed the experiments; J.S. and J.P. performed the experiments; J.H. and M.Z.A.B. analyzed the data; J.H. and H.O. contributed reagents/materials/analysis tools; J.S. and K.K. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Nam, J.; Seo, W.; Bae, J.; Cho, Y. Design and development of a u-Health system based on the ISO/IEEE 11073 PHD standards. In Proceedings of the Asia-Pacific Conference on Communications, Sabah, Malaysia, 2–5 October 2011; pp. 789–793.
2. Ramesh, M.; Anand, S.; Rekha, P. A mobile software for health professionals to monitor remote patients. In Proceedings of the International Conference on Wireless and Optical Communications Networks, Indore, India, 20–22 September 2012; pp. 1–4.
3. De Werf, F.; Bax, J.; Betriu, A.; Blomstrom-Lundqvist, C.; Crea, F.; Falk, V.; Filippatos, G.; Fox, K.; Huber, K.; Kastrati, A.; et al. Management of acute myocardial infarction in patients presenting with persistent ST-segment elevation: The task force on the management of ST-segment elevation acute myocardial infarction of the European Society of Cardiology. *Eur. Heart J.* **2008**, *29*, 2909–2945.

4. Chan, C.; Chou, W.; Chen, C.; Ho, Y.; Lin, Y.; Ma, H. Energy efficient diagnostic grade mobile ECG monitoring. In Proceedings of the IEEE International New Circuits and Systems Conference, Montreal, QC, Canada, 17–20 June 2012; p. 153.
5. Le, H.; Kuttel, M.; Chandran, G. An electronic health care—Cardiac monitoring system. In Proceedings of the IEEE International Conference on Communications Workshops, Cape Town, South Africa, 23–27 May 2010; pp. 1–5.
6. Lee, R.; Chen, K.; Hsiao, C.; Tseng, C. A Mobile Care System With Alert Mechanism. *IEEE Trans. Inf. Technol. Biomed.* **2007**, *11*, 507–517.
7. Pan, J.; Tompkins, W. A Real-time QRS Detection Algorithm. *IEEE Trans. Biomed. Eng.* **1985**, *3*, 230–236.
8. Quinlan, J. *C4.5: Program for Machine Learning*; Morgan Kaufman Publisher: San Francisco, CA, USA, 1993.
9. Mitchell, T. *Machine Learning*; McGraw-Hill International Editions; McGraw-Hill Inc.: New York, NY, USA, 1997.
10. Witten, I.; Frank, E. *Practical Machine Learning Tools and Techniques*; Morgan Kaufmann: Burlington, MA, USA, 2005.
11. Rindfleisch, T. Privacy, information technology, and health care. *Commun. ACM* **1997**, *40*, 92–100.
12. Hodge, J.G. Health information privacy and public health. *J. Law Med. Ethics* **2007**, *31*, 663–671.
13. Barrows, R.C.; Clayton, P. Privacy, confidentiality, and electronic medical records. *J. Am. Med. Inf. Assoc.* **1996**, *3*, 139–148.
14. Moody, G.; Mark, R. The impact of the MIT-BIH Arrhythmia Database. *IEEE Eng. Med. Biol. Mag.* **2001**, *20*, 45–50.
15. Simunic, D.; Tomac, S.; Vrdoljak, I. Wireless ECG monitoring system. In Proceedings of the International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronics Systems Technology, Aalborg, Denmark, 17–20 May 2009; pp. 73–76.
16. Yu, B.; Xu, L.; Li, Y. Bluetooth low energy (BLE) based mobile electrocardiogram monitoring system. In Proceedings of the International Conference on Information and Automation, Shenyang, China, 6–8 June 2012; pp. 763–767.
17. Abrar, S.; Aziz, U.S.; Choudhry, F.; Mansoor, A. Design and implementation of an embedded system for transmitting human ECG and web server for emergency services and remote health monitoring: A low cost ECG signal simulator and its transmitter, to send and store data in electronic databases, in remote location, to be accessed by authorized personnel when needed. In Proceedings of the International Conference on Open Source Systems and Technologies, Lahore, Pakistan, 20–22 December 2012; pp. 1–6.
18. Tello, P.; Manjarres, O.; Quijano, M.; Blanco, A. Remote monitoring system of ECG and temperature signals using Bluetooth. In Proceedings of the International Symposium on Information Technology in Medicine and Education, Hokkaido, Japan, 3–5 August 2012; pp. 860–863.
19. Shihu, W.; Fude, W. A remote online monitoring system for dam. In Proceedings of the International Conference on Instrumentation, Measurement, Computer, Communication and Control, Beijing, China, 21–23 October 2011; pp. 1000–1002.
20. Ko, S.; Wang, K.; Lian, W.; Kao, C. A portable ECG recorder. In Proceedings of the International Conference on Consumer Electronics, Communications and Networks, Yichang, China, 21–23 April 2012; pp. 3063–3067.
21. Ma, T.; Shrestha, P.; Hempel, M.; Peng, D.; Sharif, H.; Chen, H. Assurance of Energy Efficiency and Data Security for ECG Transmission in BASNs. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 1041–1048.
22. Lu, R.; Lin, X.; Shen, X. SPOC: A Secure and Privacy-Preserving Opportunistic Computing Framework for Mobile-Healthcare Emergency. *IEEE Trans. Parallel Distrib. Syst.* **2013**, *24*, 614–624.
23. Peter, S.; Reddy, B.; Momtaz, F.; Givargis, T. Design of Secure ECG-Based Biometric Authentication in Body Area Sensor Networks. *Sensors* **2016**, *16*, 570.
24. Tan, R.; Perkowski, M. Toward Improving Electrocardiogram (ECG) Biometric Verification using Mobile Sensors: A Two-Stage Classifier Approach. *Sensors* **2017**, *17*, 410.
25. Park, J.; Lee, K.; Kang, K. Arrhythmia Detection from Heartbeat using k -Nearest Neighbor Classifier. In Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine Workshops, Shanghai, China, 18–21 December 2013; pp. 15–22.
26. Haataja, K.; Toivanen, P. Two practical man-in-the-middle attacks on bluetooth secure simple pairing and countermeasures. *IEEE Trans. Wirel. Commun.* **2010**, *9*, 384–392.

27. Minar, N.; Tarique, M. Bluetooth security threats and solutions: A survey. *Int. J. Distrib. Parallel Syst.* **2012**, *3*, 127–148.
28. Pfitzmann, A.; Köhntopp, M. Anonymity, unobservability, and pseudonymity: A proposal for terminology. In *Designing Privacy Enhancing Technologies*; Springer: Berlin/Heidelberg, Germany, 2009.
29. Rao, J.; Rohatgi, P. Can pseudonyms really guarantee privacy? In Proceedings of the 9th USENIX Security Symposium, Denver, CO, USA, 14–17 August 2000; pp. 14–17.
30. Clayton, R.; Danezis, G.; Kuhn, M. Real World Patterns of Failure in Anonymity Systems. In *Information Hiding: Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2001; Volume 2137, pp. 230–245.
31. Han, S.; Liu, V.; Pu, Q.; Peter, S.; Anderson, T.; Krishnamurthy, A.; Wetherall, D. Expressive privacy control with pseudonyms. In Proceedings of the ACM SIGCOMM, Hong Kong, China, 12–16 August 2013; pp. 291–302.
32. Glodwasser, S.; Micali, S. Probabilistic encryption & how to play mental poker keeping secret all partial informatio. In Proceedings of the 14th Annual ACM Symposium on Theory of Computing, San Francisco, CA, USA, 5–7 May 1982; pp. 365–377.
33. Prakash, E.; Madanmohan. How to tell heart rate from an ECG? *Am. Physiol. Soc.* **2005**, *29*, 57.
34. Rodriguez, J.; Goni, A.; Illarramendi, A. Real-time classification of ECGs on a PDA. *IEEE Trans. Inf. Technol. Biomed.* **2005**, *9*, 23–34.
35. Quinlan, J. A Real-time QRS Detection Algorithm. *Mach. Learn.* **1986**, *1*, 81–106.
36. Goldreich, O. *The Foundations of Cryptography, Vol. 2, Basic Applications*; Cambridge University Press: New York, NY, USA, 2004.
37. Goldreich, O. *The Foundations of Cryptography, Vol. 1, Basic Tools*; Cambridge University Press: New York, NY, USA, 2001.
38. Fluhrer, S.; Mantin, I.; Shamir, A. Weaknesses in the key scheduling algorithm of RC4. In Proceedings of the 8th Annul International Workshop on Selected Areas in Cryptography, Toronto, ON, Canada, 16–17 August 2001; pp. 1–24.
39. Vohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the International Joint Conference on Artificial Intelligence, Montreal, QC, Canada, 20–25 August 1995; pp. 1137–1143.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).